Bulletin No. 166

1927

Agricultural Research Institute, Pusa

Sampling for Rice Yield in Bihar and Orissa

BY

J. A. HUBBACK, I.C.S.

CALCUTTA : GOVERNMENT OF INDIA
CENTRAL PUBLICATION BRANCH
1927

Government of India Publications are obtainable from the Goevrnment of India Central Publication Branch, 3, Government Place, West, Calcutta, and from the following Agents :—

EUROPE.

OFFICE OF THE HIGH COMMISSIONER FOR INDIA,
42, GROSVENOR GARDENS, LONDON, S.W. 1.
And at all Booksellers.

INDIA AND CEYLON.

Provincial Book Depôts:

MADRAS :—Office of the Superintendent, Government Press, Mount Road, Madras.
BOMBAY :—Superintendent, Government Book Depôt, Town Hall, Bombay.
SIND :—Library attached to the Office of the Commissioner in Sind, Karachi.
BENGAL :—Office of the Bengal Secretariat Book Depôt, Writers' Buildings, Room No. 1, Ground Floor, Calcutta.
UNITED PROVINCES OF AGRA AND OUDH :—Office of the Superintendent of Government Press, United Provinces of Agra and Oudh, Allahabad.
PUNJAB :—Office of the Superintendent, Government Printing, Punjab, Lahore.
BURMA :—Office of the Superintendent, Government Printing, Burma, Rangoon.
CENTRAL PROVINCES AND BERAR :—Office of the Central Provinces Secretariat, Nagpur.
ASSAM :—Office of the Superintendent, Assam Secretariat Press, Shillong.
BIHAR AND ORISSA :—Office of the Superintendent, Government Printing, Bihar and Orissa, P. O., Gulzarbagh, Patna.
COORG :—Office of the Chief Commissioner of Coorg, Bangalore.
NORTH-WEST FRONTIER PROVINCE :—Office of the Manager, Government Printing and Stationery, Peshawar.

Thacker, Spink & Co., Calcutta and Simla.
W. Newman & Co., Ltd., Calcutta.
R. Cambray & Co., Calcutta.
S. K. Lahiri & Co., Calcutta.
The Indian School Supply Depôt, 309, Bow Bazar Street, Calcutta, and 226, Nawabpur, Dacca.
Butterworth & Co. (India), Ltd., Calcutta.
Rai M. C. Sarcar Bahadur & Sons, 90-2A, Harrison Road, Calcutta.
The Weldon Library, 17, Park Street, Calcutta.
Standard Literature Company, Limited, Calcutta.
Association Press, Calcutta.
Chukervertty, Chatterjee & Co., Ltd., 13, College Square, Calcutta.
The Book Company, Calcutta.
James Murray & Co., 12 Government Place, Calcutta (for Meteorological Publications only)
Ray Choudhury & Co., 68-5, Asutosh Mukherjei Road, Calcutta.
The Oriental Publishing House, 34, Cornwallis Street, Calcutta.
B. C. Basak Esq., Proprietor, Albert Library, Dacca.
Higginbothams, Madras.
V. Kalyanarama Iyer & Co., Madras.
P. R. Rama Iyer & Co., Madras.
Rochouse and Sons, Madras.
G. A. Natesan & Co., Publishers, George Town, Madras.
The Modern Stores, Salem, Madras.
Theosophical Publishing House, Adyar, Madras.
Bright & Co., Trivandrum.
The Booklover's Resort, Taikad, Trivandrum, South India.
V. S. Swaminathan, Bookseller, West Tower Street, Madura.
E. M. Gopalakrishna Kone, Pudumandapam, Madura.
Vijapur & Co., Vizagapatam.
Thacker & Co., Ltd., Bombay.
D.B. Taraporevala, Sons & Co., Bombay.
Sunder Pandurang, Bombay.
Ram Chandra Govind & Sons, Kalbadevi, Road, Bombay.
N. M. Tripathi & Co., Booksellers, Princes Street, Kalbadevi Road, Bombay.
Proprietor, New Kitabkhana, Poona.
The Manager, Oriental Book Supplying Agency, 15, Shukrawar, Poona City.
R. S. Gondhalekar's Book Depôt, Publisher and Bookseller, Budhwar Chawk, Poona City.
Managing Director, Co-operative Bookstall, Booksellers and Publishers, Poona City.
Rama Krishna Bros., Opposite Vishrambag, Poona City.
Karsandas Narandas & Sons, Surat.
Mangaldas & Sons, Booksellers and Publishers, Bhaga Talao, Surat.
Mrs. Radhabai Atmaram Sagoon, Kalbadevi Road, Bombay.
A. H. Wheeler & Co., Allahabad, Calcutta and Bombay.
R. B. Umadikar & Co., the Bharat Book Depot, Dharwar.
The Standard Bookstall, Karachi, Quetta, Delhi, Murree and Rawalpindi.
The Karachi Book Depot, Elphinstone Street, Camp, Karachi.
The Standard Bookstall, Quetta.
U. P. Malhotra & Co., Quetta.
J. Ray & Sons, 43, K. & L. Edwardes Road, Rawalpindi and Murree.
The Standard Book Depôt, Lahore, Lucknow, Nainital, Mussoorie Dalhousie, Ambala Cantonment and Delhi.
N. B. Mathur, Supdt., Nazir Kanun Hind Press, Allahabad.
The North India Christian Tract and Book Society, 18, Clive Road, Allahabad.
Ram Dayal Agarwala, 184, Katra, Allahabad.
Manager, Newal Kishore Press, Lucknow.
The Upper India Publishing House, Ltd., 41, Aminabad Park, Lucknow.
Munshi Seeta Ram, Managing Proprietor, Indian Army Book Depôt, Juhi, Cawnpore.
Rai Sahib M. Gulab Singh & Sons, Mufid-i-Am Press, Lahore and Allahabad.
Rama Krishna & Sons, Booksellers, Anarkali, Lahore.
Puri Brothers, Booksellers and Publishers, Katcheri Road, Lahore.
The Tilak School Bookshop, Lahore.
The Standard Bookstall, Lahore.
Manager of the Imperial Book Depôt, 63, Chandney Chawk Street, Delhi.
Oxford Book and Stationery Company, Delhi.
Supdt., American Baptist Mission Press, Rangoon.
Proprietor, Rangoon Times Press, Rangoon.
The Modern Publishing House, Ltd., 30, Phayre Street, Rangoon.
The International Buddhist Book Depôt, Post Box No. 971, Rangoon.
Burma Book Club, Ltd., Rangoon.
Manager, the "Hitavada," Nagpur.
Bhisey Brothers, Booksellers and Stationers, Sitabaldi, Nagpur.
S. C. Talukdar, Proprietor, Students & Co., Cooch Behar.
Times of Ceylon Co., Ltd., Colombo.
The Manager, Ceylon Observer, Colombo.
The Manager, The Indian Book Shop, Benares City.
The Srivilliputtur Co-operative Trading Union, Ltd., Srivilliputtur (Satur S. I. R.).
Banwari Lal, Esq., Pakariya Street, Pilibhit, United Provinces.
The Manager, Educational Book Depot, Jubbulpore.
Raghunath Prasad & Sons, Patna City.
Dandekar, Brothers, Indore City.
The Hyderabad Book Depot, Chaderghat, Hyderabad (Deccan).

Sampling for Rice* Yield in Bihar and Orissa.

(*Received for publication on 10th November 1926.*)

The need for reasonably accurate statistics of the human population of a country has long been recognized throughout the civilized world. The idea that statistics of the production which supports that population are of well-nigh equal importance is rapidly gaining general recognition. It is, therefore, perhaps superfluous to argue in its favour, but it may be useful to indicate what objects a well-founded estimate of the yield of rice can fulfil.

In the first place, since rice occupies a little more than half of the cultivated area in the province of Bihar and Orissa, employs 4-5th of the total population for six months of the year and feeds the greater part of it throughout the year, its supreme importance in any survey of production can scarcely be denied. It probably provides at least one-half of the rural contribution to the "provincial dividend", and that contribution is still vastly more important than that of the urban and industrial workers. Hence if a survey of production is desirable, good statistics of rice yield are essential.

Apart from the uses to which those who direct or influence the policy of the State may put the statistics of production, there is no doubt that traders want information of crops as they come forward. Good information enables them to operate with more confidence and not only saves them from alternating between prosperity and financial disaster, but greatly benefits the public by stabilizing prices. For this purpose it is not enough to know what is the production of the whole province, but knowledge of the conditions of the districts or even smaller units is very valuable. Lastly, reasonably accurate knowledge of the real normal yields of important crops even for fairly large tracts would give guidance to courts who are required to estimate the value of produce for the decision of rent suits, for commuting produce rent or for settling fair cash rents.

---

* I have used the word "rice" throughout as the ordinary description of the crop. As a matter of fact the samples taken are, strictly, "paddy" *i.e.* unhusked rice.

Hitherto lack of such knowledge has frequently led to considerable injustice.

The statistics of area under the rice crop are claimed to be fairly accurate. They are founded ultimately on the figures obtained from detailed measurement in a particular year on the ground. No doubt new land has been brought under rice in some places, the year of survey may have been a particularly favourable or unfavourable one, or to a limited extent other crops may have replaced rice. Allowance is made for these factors year by year, and perhaps there is no very large error. The question is examined further in the latter part of this bulletin. But in the figure for yield there is certainly room for bad mistakes. This has been fixed under the advice of the Director of Agriculture at a definite figure as normal, and is varied year by year by applying for each district a percentage, indicating how far the crop of the year falls below or rises above the normal. Now in the first place it is exceedingly probable that the normal rice yield differs materially between districts. But, apart from that, there is no machinery for fixing with any reasonable accuracy the percentage of the year. All that is done at present is for the local police officers to make a guess, at which in succession the Subdivisional Officer, the District Officer, and the Director of Agriculture guess again. When it is considered that the percentage depends ultimately on the effect of the weather on very various soils, cultivated with varying degrees of skill and enterprise, planted with different kinds of rice, protected by irrigation works of greater or less efficiency or completely unprotected, liable to or immune from crop pests, and finally harvested over a period of nearly three months, it becomes apparent that the guessing ability of the officers concerned has to be remarkable. Unfortunately, too, not one of them has the least chance of finding out whether his guess was fairly right or wildly wrong. Hence the existing statistics of rice production are, I believe, the result of applying to a fairly accurate figure of area, an arbitrary standard of normal yield, and a pure guess of the condition of the year. I am convinced that by an expenditure of a sum of money, that is not excessive in view of the importance of the matter, the last two and perhaps the first could be replaced by a scientific estimate, in which the possibilities of error can be gauged. Incidentally the scheme proposed, though it would not completely solve the problems of the courts to which reference has been made, would go a great way to do so.

I return to the point of the normal yield. This is based on a number of crop-cutting experiments made at different times in different places,

and it is to the defects of these experiments, with the consequent uncertainty of the conclusions based on them, that I now wish to draw attention.

The commonest method of estimating yield, apart from pure guessing from the appearance of the field at harvest time, is that described in the standard account of the *danabandi* (appraisement) system of rent payment. It consists of the tenant cutting a small area where the crop is thinnest and the landlord where the crop is heaviest, mixing the two together and taking the result as a fair sample. To secure accurate results this presupposes that the true mean lies midway between the highest and lowest sample, a condition which is very rarely found in practice. The results so obtained are almost always far too high, *i.e.*, unfair to the tenant. In an acre field of rice bearing a mean yield of 15 maunds it might well be possible to find an area of 1-20th of an acre yielding 40 maunds an acre, but it would be obviously impossible to find another bearing ***minus*** 10 maunds an acre. The method is not far different from attempting to estimate the average income of a population by taking half the sum of the incomes of the poorest man and the richest man in it.

A second method is that at present prescribed. An officer of some standing visits the tract, of whose yield he is required to form an estimate, and selects a fairly large area, usually 1-10th of an acre, as containing an average crop. This he cuts and carries and in due course threshes and weighs. If he has time he makes two or more experiments, but it is obvious that he cannot usually manage to make a large number of such experiments. There are two main objections to this method: the first is that it depends entirely for its accuracy on the ability of the officer to select. He has to make his selection from a large number of fields growing different varieties at different points of maturity. It is very easy to be misled as to yield by the strength of straw, heavy straw being by no means always correlated with heavy yield of grain. Again it is difficult to give proper weight to the fields which will give little or no yield at all, and even in a normal year such fields are by no means rare. These factors make accurate selection very difficult. The second objection is even more important, *viz.*, that there is no possible way of estimating what is the probability that the result of such selections is within a given range from the true mean yield. The method is comparable to estimating the average income of the population of a town by watching the streets for a few days and then picking out a man, who looked to be in average circumstances and discovering what his income is.

The only way in which a satisfactory estimate can be formed is by as close an approximation to random sampling as the circumstances permit, since that not only gets rid of the personal element of the experimenter, but also makes it possible to say what is the probability that the result of a given number of samples will be within a given range from the true mean. To put this in definite language, it should be possible to find out how many samples will be required to secure that the odds are at least 20 to 1 on the mean of the samples being within one maund of the true mean.

This is the problem to be solved. Before describing the progress made towards a solution, it is desirable to explain some terms used in modern mathematical statistics, on which the solution depends. These will, no doubt, be quite familiar to many of my readers, to whom I apologize for the digression. The first term is the arithmetical " mean ", which is the real object of our search. It is, of course, the sum of all the individual results of the cuts or samples made divided by the number of cuts or samples. The true " mean " is the figure which would be obtained if the number of samples were increased, so as to cover the whole subject matter of sampling exhaustively. The whole subject matter of sampling is known as the " population ", viewed as made up of a very large number of samples with varying values. The manner in which the values of the samples are numerically arranged is called the " distribution ". Clearly the distribution may in some populations be very much wider than in others. The number of samples having a given value, or having a value lying within given limits, is called the " frequency " for that value or for the range of those limits. The range of limits is known as the " class interval ", and, for convenience of calculation, values of samples, when these vary continuously, are lumped together in equal class intervals and treated as if they all fell at the middle point of their class interval. A most important term is the " standard deviation " of the distribution. This is a measure of the amount of dispersion or " scatter " about a particular value, usually about the mean. It is a perfectly definite value expressed in the same unit as that in which the value of each sample is expressed. It is obtained as follows. The value of each sample differs by a definite figure from the particular value, about which the dispersion is sought, *e.g.* from the mean. This difference is squared and (where the samples have been lumped together in class intervals) multiplied by the frequency of the class interval. The results are added together and divided by the sum of all the frequencies, and the square root of that result is the standard deviation. A simple illustration

will explain the process. Suppose that 8 samples give the following results.

Frequency	Value	Difference from the mean which is 6	Square of the difference multiplied by the frequency
1	4	−2	4
4	5	−1	4
1	6	0	0
0	7	+1	0
0	8	+2	0
2	9	+3	18
TOTAL 8			TOTAL 26

The standard deviation is the square root of 26 divided by 8 or 1·8 nearly.

There remains to be explained yet one more term, "the normal curve of error". This is a graphic representation of a certain class of frequency distribution, which has, among other properties, that of symmetry about the mean. To this class many of the distributions actually found in practice more or less closely conform. It has the very great advantage that the chance that a particular sample will fall within a definite range on either side of the mean can be calculated at once from standard tables, and this is what makes it so widely used in statistical practice. The properties of other classes of frequency distributions have been investigated, but their use involves difficult and laborious calculation, which makes them unsuitable for many kinds of work. It is not infrequently assumed that the determination of the standard deviation of any frequency distribution at once enables one to say what are the odds that the mean of sampling falls within given limits on either side of the true mean. This is untrue, and it is always necessary to see that the distribution under scrutiny conforms closely to the normal curve of error, before making deductions, which are true for that distribution but not for others.

Evidently then, before a practical method of obtaining a satisfactory value for the mean yield of the rice crop from a system of sampling could be recommended, it was necessary to get a clear idea of the distributions, which random sampling will give. The first attempt was made under the direction of Mr. Dobbs, Director of Agriculture, Bihar and Orissa, in 1921 in the Nawada thana of Gaya. His method was

to divide up the map into equal squares, in order to pick out the villages in which samples were to be taken, and then to sample the fields bearing the survey number 50, or the nearest fields planted with rice. For this method it was necessary to find out the date on which the crop on each field would be ready for harvest and to arrange to visit the village on that date. The practical difficulties involved made me search for another method, more automatic in the distribution of sampling in time and space. It is unfortunately impossible to use the result of Mr. Dobbs' experiment, because three samples were taken in each field and the results lumped together. This has disguised the frequency distribution so much as to make deductions extremely hazardous.

In 1923 with the help of Mr. Davies, the Settlement Officer of the Santal Parganas district, 400 random samples were obtained in a part of Godda thana. The sampler spent from the 4th December to the 22nd December travelling from village to village through the area comprising about 100 square miles. Wherever he found, on the day of his visit, the crop being actually harvested, he took a sample. In the following year the same method was applied under the supervision of the Settlement Officers, Messrs. Davies, Mansfield and Gokhale, in the Godda thana of Santal Parganas, the Jajpur thana of Cuttack, and the Purulia thana of Manbhum. The only difference was that four samples instead of one were taken, where practicable, on each field in which the harvest was in progress.

In 1925 I obtained the permission of the Bihar and Orissa Government to continue and expand the experiment in Santal Parganas and Orissa. An officer was employed in each of eight subdivisions, averaging nearly 1,000 square miles in extent, and in each some 15 centres were taken, so as to cover as widely as possible the subdivisional area. At each of these the officer was to spend four days, moving out on the first day two miles north, turning to the right and returning from the north-east, and on the succeeding days traversing the other three quadrants in the same manner. The aim was to distribute the sampling as regularly as possible both *in time* throughout the period of harvest and *in space* over the area sampled. The limitation in distance to be traversed secured that the number of samples obtained would be few in tracts where harvesting was sporadic and many in tracts where it was in full swing, thus correcting any tendency to over-emphasize the yields from relatively unimportant types. The instruction to take four samples from each field was withdrawn for reasons given below. For the supervision of this year's sampling I am indebted to Messrs. Mansfield and Houlton.

In all three years the actual process of taking each sample has been the same. The apparatus consists of a small solid block of wood, on which are fixed four pairs of guides defining two lines at an angle of 60 to one another. Along these lines through the guides are pushed two narrow wooden battens, which pass among the stalks of the paddy. Each bears a stop, so that, when this rests against the first pair of guides the length from the point where the battens cross to the slots near the ends of the battens is $67\frac{1}{4}$ inches. From one of these slots a third batten is pushed through the stalks to the slot on the other batten, and, when this third batten is fixed in the slots, an equilateral triangle is formed comprising an area of 1/3,200 of an acre. The paddy stalks just outside the battens are bent back, and the crop inside cut and threshed out there and then. The grain is placed in a numbered bag and dried for three or four days and then weighed. The size of the triangle has been chosen to give one maund an acre for every standard tola of the sample, while the triangular shape is intended to avoid the error caused by either including or excluding a whole line of the planted paddy. The weight can be accurately determined to half a tola in ordinary country scales by double weighing. The sampler is instructed to go to the centre of the *ail* (field boundary) roughly parallel to the line on which the harvest is proceeding and take three paces into the field before laying down the solid block of wood which guides the battens. He is warned against the mistake of putting the block down so that the triangle of battens encloses it instead of excluding it, a mistake to which beginners are liable.

The results obtained in the last three years are given in Tables I and II. Besides those, to which reference has already been made, there are a set of samples taken in 1925 on the Kanke Agricultural Farm in Ranchi district and a set taken in the Atri thana of Gaya district in the same season. The system has thus been tried in the Orissa deltaic tract, on the Chota Nagpur plateau of which Ranchi forms a part, in the broken country of Santal Parganas and Manbhum, and in the alluvial Gangetic plain of North Santal Parganas and Gaya. It has been put in practice in tracts, where the harvest has been plentiful and in others, where it has been scanty. In 1925 the sampling extended in seven tracts over the whole period of the rice harvest, excluding only the very early *bhadoi* and the summer rice, reaped in March, which are harvests of very little importance. Conclusions that can be drawn from the figures may, therefore, be applied with safety to the whole province except perhaps the North-Gangetic tract.

I have relegated to the Appendix the technical discussion of the validity of my conclusions, which will be appreciated only by those readers, who have made a somewhat extensive study of statistical theory. It is enough to say here that the mean yield on the harvested area can, I think, be determined for a tract of about 1,000 square miles with an accuracy such that it will be out by more than one maund per acre in only one case out of twenty. It will be very seldom out by more than a maund and a half. To secure such determination, it is necessary to fix 12 centres, spread as evenly as possible over the area, and to put down against each centre the day on which sampling is to be done. These days should be spread evenly over the period of harvest. The sampler should go out a fixed distance in one direction and, circling round, return from another direction, so that it is secured that he covers approximately the same area on each day. He should cut one sample from each field, where he finds harvesting in progress. He will be able ordinarily to get from 30 to 40 samples in a day when the harvest is in full swing and 10 to 20 when it is slack. It is certainly not impossible to supply samplers for this work, which is extremely simple once it is explained. The men employed hitherto have been young men fresh from college. Any gazetted officer and most other officers could do it quite well. In fact, I have found that peons attached to the sampling officers fully understand what has to be done after a week or so. It is not in the least essential that the same officer should carry out all the sampling of a tract.

The method is applicable to small or large tracts. The same number of centres will probably give a slightly greater degree of accuracy for a smaller and a slightly less degree for a larger tract, because the standard deviation on the whole increases with the size of the tract, though not at all rapidly. Thus the standard deviation of the whole of the Santal Parganas is only 9·26 maunds, actually less than that of Deoghar, though greater than that of any other subdivision. To determine the mean yield of a village or an agricultural farm it would probably be advisable to make at least 10 cuts on each of 10 days. It would be better to make at least 20 cuts. (It may be noted that the standard deviation of Kanke in 1925 was nearly the highest figure found so far.)

The method could also be applied to other crops, especially cereals. It would be more difficult to apply it to crops, for which the marketable produce is obtained by processes more elaborate than threshing, or is bulky and awkward to weigh, such as jute or sugarcane. It is also an advantage if the crop is common enough to enable a fair number of samples to be taken on one day.

It is no advantage to take a large number of samples from places very close together, where the crops will naturally be very much the same on the same day. The degree of accuracy is not seriously improved by such practice. This explains why there is no need to take large samples instead of the handy samples obtained by my method. A sample of one-tenth of an acre is merely 320 of my samples taken in juxtaposition. It simply gives a determination of the mean yield of that particular field, which is not more effectively accurate than that given by say four small samples. Even four samples instead of one are not worth while, because in the great majority of cases they do not differ among themselves enough to affect the mean or the standard deviation of the whole set of samples. This may be illustrated from the columns for "all classes" and "1st cutting" for Santal Parganas, Godda thana 1924. Technically speaking, there is very high correlation between the individuals of such groups of samples, which makes the ordinary rule, that the standard deviation of the mean is the standard deviation of the population divided by the square root of the number of samples, quite inapplicable.

A feature of all the frequency distributions obtained is their tendency to have a fairly level crest, the frequency not varying greatly for some four or five class intervals around the mean. This emphasizes the difficulty of "selecting an average field", the present method of experimenting. There will be nearly as many places, where the crop is 4 maunds above or below the mean, as there will be where it is within one maund of it. Thus from the very nature of the rice crop, apart from the physical impossibility of surveying the whole of it simultaneously or nearly so, and the personal error of the experimenter, the present method of estimating is likely to be very far from accurate.

My method gives a reasonably close approximation for the mean yield of the *harvested area*. Consequently to get the gross yield it is necessary to determine what is that harvested area. It is evidently not the area planted or sown, since in practically all years some fields bear no crop worth cutting. It also excludes the field ridges (*ails*), which are usually measured in with the fields at a cadastral survey. Since the cadastral survey figure is the basis for the cropped area of the year, this second point is of importance. No very satisfactory determination of the area occupied by ridges has yet been made, but it would appear to vary between nearly 10 per cent. in hilly country to scarcely more than 2 per cent. in level country. It is usually taken as 5 per cent.

It is much more difficult to say what is the range of variation between the planted (or sown) area and the harvested area. In a bad year, when the rains stop early in September, it may well be 50 per cent, all over a subdivision. The planted area will also vary from year to year. At present we depend on estimates, which are originally those of the village chaukidar, though they are much more capable of intelligent check by higher authorities than estimates of yield are. It is possible, for instance, by driving through a tract in a motor car to get some idea of the extent to which the rice crop is going to fail altogether. It is usually believed that the estimate of cropped area is sufficiently accurate. It may be, but the belief is intuitive rather than rational.

It would certainly be a great advantage if the harvested area could be obtained by some form of sampling. A method was tried in 1925 in Santal Parganas. The results, if not very encouraging, are interesting. The plan was to make the sampler march from centre to centre across country as near as he conveniently could in a straight line. After certain intervals of time he had to count 100 paces and note how many of these ended on harvested rice land. (There is no serious difficulty in doing this in January and early February, since little rice land is reploughed by then.) Each count of 100 would give a definite percentage, in many cases 0, in many nearly 100, the difference from the full 100 being due to the presence of field ridges. The mean of these percentages would give the percentage of harvested rice land in the total area of the tract sampled. Six men were employed for nearly a month each and between them made 1971 counts. From these a mean of 47·22 per cent. was obtained. The standard deviation was 37·10. This means that it is about 21 to 1 that the true percentage is between 49·4 and 45. The accuracy obtained is hardly sufficient to justify the employment of six men for a month, in each district. Further the work is strenuous and tedious, and it is probable that it would be in practice shirked, and results fudged. The difficulty of excluding large patches of thick jungle or other inaccessible country both from the sampling and the sampled area has yet to be overcome. But it is still possible that some method on similar lines may prove practicable.

There is just one other caution which I would address to persons who have to estimate gross yields. That is that yields of rice are ultimately based, however imperfectly at present, on yields of paddy. A standard figure for the loss of weight by husking is taken, usually one-third. It is probable that this varies a good deal with the kind of paddy, and there is good reason to believe that in Orissa the loss is considerably more than one-third. It would be well if the Agricultural Department

investigated this point on their farms. If any kind of paddy husks out much below the usual, it is clearly not such a good economic proposition, as an excess of unhusked yield might suggest. Possibly the plant breeders may be able to produce a paddy, which loses less than one-third in husking, and, if so, the benefit might well be as great as that obtained from increased yield.

My general conclusions are that sampling by my method on 12 days throughout the harvest from different centres would secure a figure for the mean yield of a harvested area of 1,000 square miles correct within one maund in about 95 per cent. of cases, and that with the existing revenue and agricultural staff there would be no serious difficulty in substituting this method for the unscientific method of sampling now prescribed.

To the officers I have mentioned, under whose direction the samples were taken, and to the staff employed by them on the work I offer my cordial thanks, as I do to Mr. Dobbs for his advice and encouragement.

APPENDIX.

It is often assumed that if a number (n) of samples are taken and the standard deviation of the whole population (σ_0) is known, then the standard deviation (σ_m) of the mean of n samples is given by $\sigma_m^2 = \sigma_0^2 n$. This however, presupposes (1) that the samples are taken by the process of "simple" or "random" sampling, *i.e.*, that every portion of the population is equally likely to be sampled every time, and (2) that the distribution of the population is such that the distribution of the means of n samples closely approximates to the "normal" curve of error.

I have attempted to investigate both these points before recommending the method already tried for general application. The application of "simple" sampling to the rice crop would involve picking out the places for sampling by some random method such as putting numbered cards representing all the rice plots of the area in a box and drawing out the required number of cards, and then arranging to visit the plots at the time when the crop is ripe. Such a method is, I consider, impracticable. My method is practicable but is not "simple" sampling. There is some degree of selection. It has to be seen whether the adoption of it in place of "simple" sampling increases or decreases the rate at which the standard deviation of the mean diminishes as the number of cuts rises.

If an equal number of cuts were taken on each day, say 20, the process would be equivalent to (1) taking one cut from each separate tract to make up a sample, and then (2) taking the mean of 20 such samples; the "tract" being the country which the sampling officer can cover from a centre. Process (1) theoretically gives a set of samples with means having a standard deviation given by $\sigma_{m_1}^2 = \frac{\sigma_0^2}{n_1} - \frac{s_m^2}{n_1}$ where σ_0 is again the standard deviation of the population, s_m that of the mean values of the tracts and n_1 the number of the tracts. Process (2) is simple sampling of the means of the samples obtained by process (1), and will give a mean with a standard deviation given by $\sigma_{m_k}^2 = \frac{\sigma_{m_1}^2}{k}$, where k is the number of samples grouped together (in this case 20).

To test the theory I have taken from the original record of Rajmahal 1925 the first eight cuts of each day, writing the 32 cuts of the first four days in one row, the 32 cuts of the second four days in the next row, and so on. (In a very few cases, where less than eight cuts were taken in a day, I have borrowed from the last cuts of the previous day or the following

day.) I have thus obtained 32 columns with 11 figures in each corresponding to the 11 centres, each surrounded by four quadrants. I have totalled the columns and taken the means. I have then taken the means of successive pairs, fours, and eights of the means so derived. I have also calculated s_m^2 by taking the means of the cuts made from each centre. I have done the same thing for Bhadrak 1925, which had 12 centres instead of 11. The results are as follows (standard deviation in maunds).

	σ_0^2	s_m^2	$\sigma_{m_1}^2$	$\sigma_{m_2}^2$	$\sigma_{m_4}^2$	$\sigma_{m_8}^2$
Rajmahal theoretical . .	69·06	34·21	3·17	1·58	·79	·40
" observed . . .	..	..	1·62	1·08	·40	·31
Bhadrak theoretical . . .	33·18	9·46	1·98	·99	·50	·25
" observed . . .	..	..	1·10	·74	·54	·11

(The theoretical value of $\sigma_{m_1}^2$ is $\sigma_0^2 - s_m^2$ divided by 11 for Rajmahal and 12 for Bhadrak. The theoretical values of $\sigma_{m_2}^2$, $\sigma_{m_4}^2$ and $\sigma_{m_8}^2$ are one half, one quarter and one-eight of $\sigma_{m_1}^2$.)

The observed results agree sufficiently well with the theory, and it seems quite clear that the standard deviation is diminished and not increased by taking samples on this plan instead of by simple sampling.

But this plan would require an equal number of cuts on each day, and this would certainly overweight the slack periods of harvest. My method avoids this, but by so doing it almost certainly increases the standard deviation of the mean as compared with that obtained by the plan just described. I cannot find any way of determining mathematically to what extent it does so, but the following test shows reasonably well what happens.

I examine below the means obtained from the cuts made to the north-east of each centre against those obtained from the cuts made to the south-east, south-west and north-west. These are given below. (Some cuts used in the general table were neglected for this purpose, because only one quadrant was sampled.)

—	Average No. of cuts in a quadrant	1st quadrant	2nd quadrant	3rd quadrant	4th quadrant	All	Standard Deviation Actual	Standard Deviation Theoretical
All S. P. .	1432¼	16·47	16·85	17·32	17·42	17·02	·25	·25
Deoghar .	280	20·01	19·12	21·31	20·23	20·15	·72	·69
Dumka .	122	19·57	20·52	19·46	20·63	20·06	·44	·71
Godda . .	342¼	12·64	14·05	15·05	15·13	14·25	·65	·35
Jamtara .	216	13·60	14·39	12·09	13·43	13·41	·78	·52
Rajmahal	312¼	21·08	21·38	22·06	22·30	21·67	·37	·47
Pakaur .	159¾	10·95	10·80	11·69	11·11	11·07	·19	·52
Bhadrak .	288¾	13·13	14·60	13·52	13·29	13·65	·49	·34

The " failures ", *i.e.*, results falling more than one maund away from the mean of the four quadrants are underlined. The actual standard deviation is that derived from the means of the four quadrants, Sheppard's correction being applied. The theoretical standard deviation is the result of dividing the standard deviation of the whole series of cuts in the area by the square root of the average number of cuts in the quadrants. The " failures " look high at first sight, but that is due to the fact that the number of cuts in a quadrant was often distinctly lower than the standard of 20 cuts at 12 centres, which would seem to be what is required to ensure a mean correct within one maund in 95 per cent. of cases. A comparison of the actual with the theoretical standard deviation shows that on the whole my method of sampling gives about as stable a mean as simple sampling. The variation in the number of cuts at the different centres about wipes out the greater stability gained by distributing the cuts to the different centres.

It is impossible to demonstrate by mathematical processes precisely what the odds are, and, without a much wider series of tests than have yet been made, empirical deductions must be somewhat uncertain. But I think it is quite safe to say that if at least 10 and not more than 40 cuts were made from each of the 12 centres at even intervals over the harvest season, the mean yield for an area of about 1000 square miles could be determined within one maund per acre nineteen times out of twenty. The " failures " would very seldom be more than $1\frac{1}{2}$ maunds out. This depends on the assumption that the frequency distributions actually found are such that the means of a large number of cuts will form a frequency distribution closely approximating to the " normal curve of error ". Where the sampling is " simple ", this is certainly true of populations following the frequency distributions obtained. If, for example, 400 samples were taken by " simple " sampling from populations like the Deoghar 1925 and the Godda 1925 harvests, the spread of the mean of the 400 samples would compare with the spread of the corresponding "normal" curve (*i.e.*, that having the same standard deviation) as follows.

Limits.	Deoghar	Normal
± ·425 maunds	1125 inside to 1000 outside	1138 inside to 1000 outside.
± ·850 "	585 " to 100 "	583 " to 100 "
±1·275 "	364 " to 10 "	330 " to 10 "
±1·700 "	342 " to 1 "	272 " to 1 "
	Godda	Normal
± ·3428 "	2517 inside to 1000 outside	2531 inside to 1000 outside.
± ·5142 "	822 " to 100 "	837 " to 100 "
± ·6856 "	3011 " to 100 "	3050 " to 100 "
± ·8570 "	1357 " to 10 "	1375 " to 10 "

It cannot then be said with certainty, what effect on the rapidity of the approach to normality would arise from the substitution of my method of sampling for "simple" sampling. But as it has little or no effect of diminishing or increasing the standard deviation of the mean, it probably has the effect of bringing the higher constants about as quickly towards their values for the normal curve. The nearest approach to a discussion of this problem is to be found in a paper by "Student" in Vol. VII pages 210 to 214 Biometrika. But the problem before him was not precisely mine, and he distinctly states that his problem is essentially a "small sample" problem. It deals with means of 25, not as mine with means of 250 or so. The higher constants for some of the distributions are given below (Professor Pearson's notation).

	β_1	β_2
Godda 1925	·527	4·116
Deoghar 1925	1·227	4·066
Pakaur 1925	1·348	6·349
Godda 1924 1st cuttings	·030	2·729
Jajpur 1924 1st cuttings	·694	3·388

Professor Bowley's tables on pages 271 and 303 of his "Elements of Statistics" enable one to show how the observed mean is likely to be about the true mean. Assuming 256 samples for each tract, 10,000 trials would be distributed thus :—

($\varkappa$ is Professor Bowley's measure of skewness and is $\mu_3 \sigma^3$ in Professor Pearson's notation.)

——	σ	$\varkappa$	Not more than 1 maund above mean.	Not more than 1 maund below mean.	More than 1 maund from mean.
Godda 1925	6·63	·64	4,888	4,956	156
Jamtara 1925	7·70	·78	4,752	4,844	404
Pakaur 1925	6·54	·86	4,900	4,982	118
Rajmahal 1925	8·31	·26	4,712	4,752	536
Deoghar 1925	11·59	1·11	3,999	4,225	1,776
Dumka 1925	8·15	·59	4,706	4,782	512
Jajpur 1925	9·09	1·51	4,536	4,714	750
Bhadrak 1925	5·76	·69	4,943	5,003	54

It may, of course, be argued that my method, even if the same or nearly the same mean were obtained in a large number of repetitions, may be defective owing to bias. That must remain a matter of opinion.

The method does, however, secure that the cutting is well distributed in space over the area to be sampled and in time over the period of the harvest. By laying down that approximately the same area shall be under sample on each of the days of sampling, it weights the importance of the contribution made to the mean by the harvest of the day. A day in the height of the harvest will contribute 30 to 40 samples; in the slacker time only 10 to 20. There is no other obvious form of bias, unless, of course, centres are selected with grave disregard to their suitability.

If it be admitted that the distribution of the means of a fairly large number of cuttings will closely approximate to a " normal curve of error " and that the standard deviation of the means is approximately that of the population divided by the square root of the number of cuttings, then the degree of accuracy can be calculated. Leaving aside the standard deviations of rice grown on particular classes of land, which are sometimes very low, it may be taken that the range of the standard deviation of rice in Bihar and Orissa is from 4 to 12 maunds. At the lower limit, 240 cuttings would give means of which only 15 out of 100,000 would be one maund or more away from the true mean. At the upper limit about 19 out of 100 would be so divergent, and about 5 out of 100 would be more than $1\frac{1}{2}$ maunds away. With a standard deviation of 8 maunds, 240 cuttings would give a mean correct within one maund about 95 times out of 100. This last is the basis of my scheme of 12 centres, from which may be expected at least 240 cuttings, but probably as many as 300. Where experience shows that the standard deviation of a tract is much higher than 8 maunds, the number of centres should be increased, if it is desired to obtain the same degree of accuracy.

It is of interest to consider what degree of accuracy would attach to a calculation of the gross yield of the whole province of Bihar and Orissa, based on the system I recommend, the harvested area being estimated as it is present. If it may be assumed that the standard deviation of the error in estimating the harvested area is about 1/20 of that estimate, it is practically certain (1,000 to 3) that the calculation of the gross yield is not more than 2 per cent. out.

TABLES

TABLE I.

Frequency distributions of rice yields, 1923-24.

Class intervals in maunds.	Santal Parganas, Godda Thana, 1923, all classes	SANTAL PARGANAS, GODDA THANA, 1924					MANBHUM, PURULIA THANA, 1924					CUTTACK, JAJPUR THANA, 1924	
		1st class lands	2nd class lands	3rd class lands	All classes	1st cutting	1st class lands	2nd class lands	3rd class lands	All	1st cutting	All	1st cutting
1	2	3	4	5	6	7	8	9	10	11	12	13	14
0—2⅛ . . .	55	12	1	26	39	8	..	1	2	3	1	28	16
2⅛—4¼ . . .	24	28	24	48	100	28	1	2	5	8	2	68	37
4¼—6⅜ . . .	44	69	49	82	200	43	6	8	40	54	18	102	45
6⅜—8½ . . .	45	128	53	90	271	71	17	11	74	102	33	124	43
8½—10⅝ . . .	71	126	86	47	259	82	18	17	84	119	37	115	42
10⅝—12¾ . . .	27	175	78	30	283	62	18	18	100	136	42	124	52
12¾—14⅞ . . .	18	223	59	60	342	79	17	31	78	126	39	83	34
14⅞—17 . . .	21	192	47	23	[illegible]	67	22	28	77	127	50	87	37
17—19⅛ . . .	19	173	27	4	204	57	31	19	70	120	40	52	14
19⅛—21¼ . . .	22	118	12	4	134	49	13	27	33	73	27	39	21
21¼—23⅜ . . .	15	51	9	3	63	17	18	16	22	56	23	24	5
23⅜—25½ . . .	14	16	6	3	25	4	16	11	18	40	20	28	11
25½—27⅝ . . .	12	4	3	..	7	1	10	8	9	27	15	21	11
27⅝—29¾ . . .	4	6	6	..	12	3	10	5	4	19	10	17	4
29¾—31⅞ . . .	1	2	2	..	4	1	7	8	2	17	10	8	5
31⅞—34 . . .	4	6	..	..	6	3	4	3	..	7	3	4	1

34—36⅛	2	..	..	..	..	..	..	..	..	..	..	2	1
36⅛—38¼	0	..	..	..	..	..	1	4	..	5	2	2	0
38¼—40⅜	1	..	..	..	..	..	..	..	..	..	..	1	1
40⅜—42½	0	..	..	..	..	..	..	..	..	..	..	1	..
42½—44⅝	1	..	..	..	..	..	..	1	..	1	..	..	..
44⅝—46¾	..	..	..	..	..	..	..	..	..	..	..	..	..
46¾—48⅞	..	..	..	..	..	..	1	..	..	1	..	..	..
Total number of cuttings.	400	1,329	462	420	2,211	575	210	218	613	1,041	372	930	380
Mean in maunds	10·67	13·83	11·74	8·56	12·70	12·63	17·76	16·37	13·40	15·59	16·03	12·25	11·81
Standard deviation in maunds.	8·06	5·30	5·38	4·71	5·60	5·68	7·48	7·50	5·36	6·59	6·91	7·13	7·20
Period of cutting	4th December to 22nd December.	14th November to 13th January.	10th November to 27th December.	10th November to 19th December.	10th November to 13th January.	10th November to 13th January.	29th November to 22nd December.	17th November to 19th December.	16th November to 13th December.	16th November to 22nd December.		17th December to 2nd January.	

The "classes" of lands are those used for rent settlement purposes, the first class being generally bottom valley lands, with a fairly secure supply of water, the third class being lands at the top of the slopes, and the second class intermediate.

TABLE II.

Frequency distribution of rice yields, 1925.

Class intervals in maunds	Godda Thana (Santal Parganas)				Jamtara Thana				Pakaur Thana				Rajmahal Thana			
	1st	2nd	3rd	All	1st	2nd	3rd	All	1st	2nd	3rd	All	1st	2nd	3rd	All
0—2⅛ . .	0	0	8	8	0	1	42	43	0	2	18	20	0	0	0	0
2⅛—4¼ . .	1	5	46	52	1	9	44	54	1	13	36	50	0	0	7	7
4¼—6⅜ . .	2	17	73	92	2	18	36	56	3	36	64	103	0	0	24	24
6⅜—8½ . .	0	42	72	114	14	41	31	86	4	42	44	90	0	1	26	27
8½—10⅝ . .	12	90	54	156	20	61	26	107	15	40	39	94	0	19	42	61
10⅝—12¾ . .	21	115	34	170	31	52	20	103	12	26	28	66	0	84	12	96
12¾—14⅞ . .	46	97	35	178	29	54	6	89	11	27	13	51	0	96	..	96
14⅞—17 . .	64	91	17	172	24	47	2	73	12	20	4	36	0	99	..	99
17—19⅛ . .	72	66	9	147	42	34	1	77	12	24	6	42	37	73	..	110
19⅛—21¼ . .	58	35	3	96	25	21	..	46	12	24	2	38	120	18	..	138
21¼—23⅜ . .	37	17	1	55	15	13	..	28	13	8	1	22	141	10	..	151
23⅜—25½ . .	23	15	..	38	18	13	..	31	5	5	0	10	133	..	..	133
25½—27⅝ . .	25	3	..	28	16	9	..	25	1	2	1	4	74	..	..	74
27⅝—29¾ . .	15	4	..	19	9	4	..	13	3	3	..	6	54	..	..	54
29¾—31⅞ . .	15	5	..	20	11	2	..	13	2	1	..	3	49	..	..	49
31⅞—34 . .	4	3	..	7	3	1	..	4	1	..	..	1	82	..	..	82
34—36⅛ . .	5	2	..	7	3	2	..	5	1	..	..	1	45	..	..	45
36⅛—38¼ . .	3	..	..	3	1	2	..	3	0	..	..	0	19	..	..	19

	2	..	..	2	0	..	..	0	0	..	..	0	23	..	..	23
40⅜—42½	..	..	..	..	2	..	..	2	0	..	..	0	10	..	..	10
42½—44⅝	..	..	..	..	1	..	..	1	0	..	..	0	3	..	..	3
44⅝—46¾	..	..	..	..	0	..	..	0	0	..	..	0	2	..	..	2
46¾—48⅞	..	..	..	..	1	..	..	1	0	..	..	0	..	..	..	..
48⅞—51	..	..	..	..	..	..	..	..	1	..	..	1	..	..	..	..
51—53⅛	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
53⅛—55¼	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
55¼—57⅜	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
57⅜—59½	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
59½—61⅝	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
61⅝—63¾	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Total number of cuttings.	405	607	352	1,364	268	384	208	860	109	273	256	638	792	400	111	1,303
Mean in maunds	19·61	14·13	8·50	14·30	18·40	13·95	5·87	13·38	16·87	12·14	7·59	11·12	26·56	14·95	7·97	21·41
Standard deviation in maunds.	5·96	5·15	4·20	6·63	7·47	6·20	3·84	7·70	7·42	6·31	4·26	6·54	4·97	2·94	2·34	8·31
Period of cutting	October 20th—January 1st				October 20th—December 17th				October 21st—December 20th				November 5th—January 2nd			

TABLE II—contd.

Frequency distribution of rice yields, 1925—contd.

Class intervals in maunds	Deoghar Thana				Dumka Thana				Jajpur Thana (Cuttack)	Bhadrak Thana	Kanke farm (Ranchi)	Atri Thana (Gaya)
	1st	2nd	3rd	All	1st	2nd	3rd	All	All	All		
0—2⅛ . .	0	0	2	2	0	0	0	0	31	2	0	6
2⅛—4¼ . .	0	1	13	14	0	0	3	3	64	36	1	8
4¼—6⅜ . .	1	4	47	52	0	1	13	14	79	63	1	15
6⅜—8½ . .	10	10	60	80	0	3	16	19	79	131	1	32
8½—10⅝ . .	16	28	58	102	0	9	16	25	79	142	2	21
10⅝—12¾ . .	25	33	42	100	5	10	13	28	68	138	2	26
12¾—14⅞ . .	17	37	19	73	8	18	18	44	52	186	4	14
14⅞—17 . .	39	40	19	98	10	22	8	49	60	171	5	49
17—19⅛ . .	32	39	21	92	30	14	20	64	41	99	6	21
19⅛—21¼ . .	48	35	14	97	20	22	7	49	39	75	3	25
21¼—23⅜ . .	31	22	10	63	25	15	5	45	29	35	3	23
23⅜—25½ . .	33	11	5	49	26	15	5	46	17	27	2	30
25½—27⅝ . .	34	10	7	51	18	6	6	30	7	18	6	23
27⅝—29¾ . .	39	14	2	55	12	0	1	13	8	12	6	2
29¾—31⅞ . .	19	5	..	24	13	0	1	14	2	5	7	14
31⅞—34 . .	26	11	..	37	16	0	0	16	6	4	4	5
34—36⅛ . .	27	4	..	31	10	2	2	14	4	2	6	3

36⅛—38¼	8	5	..	13	2	1	..	3	5	1	3	2
38¼—40⅜	16	0	..	16	2	..	..	2	5	1	1	2
40⅜—42½	15	1	..	16	3	..	..	3	3	..	2	4
42½—44⅝	6	..	..	6	5	..	..	5	0	..	1	1
44⅝—46¾	9	..	..	9	2	..	..	2	2	..	2	..
46¾—48⅞	9	..	..	9	0	..	..	0	3	..	1	..
48⅞—51	6	..	..	6	1	..	..	1	0	..	2	..
51—53⅛	5	..	..	5	..	..	..	..	2	..	0	..
53⅛—55¼	6	..	..	6	..	..	..	..	2	..	1	..
55¼—57⅜	5	..	..	5	..	..	..	..	..	..	..	..
57⅜—59½	6	..	..	6	..	..	..	..	..	..	..	..
59½—61⅝	5	..	..	5	..	..	..	..	..	..	..	..
61⅝—63¾	1	..	..	1	..	..	..	..	..	..	..	..
Total number of cuttings	494	310	319	1,123	217	138	134	489	687	1,153	72	326
Mean in maunds	27·07	18·13	11·29	20·12	24·64	18·08	14·21	19·93	12·88	13·63	26·63	17·46
Standard deviation in maunds	11·71	7·25	5·74	11·59	7·69	5·57	6·80	8·15	9·09	5·76	11·30	8·73
Period of cutting	October 20th—December 26th				October 25th—December 21st				November 23rd—December 31st.	October 23rd—January 11th.	..	December 4th—January 1st.

MGIPC—M—IV-2-59—4-7-27—750